AF Data

POLLING SUMMARY

CONTENTS

INTRODUCTION **2**

Terminology 2
Public Opinion Poll, Types of Polls & Types of Sampling 4
Polling Errors & Biases: 6

TYPES OF POLLS **8**

SMS/Online/Email 8
Interactive Voice Response (IVR) 11
Live Calling 12

POLLING AGENCIES **13**

Rasmussen 13
Quinnipiac 13
Insider Advantage 15
Trafalgar 16

DATA VIS: GRAFANA **17**

Grafana 17

Introduction

Terminology

Statistic ($\hat{p}$)

A number that can be computed from data without a parameter; estimates a parameter, but ascertained with a sample

Parameter (p)

A number that can be computed from a whole population; estimated by a statistic, but ascertained with a population

Sample (n)

A portion of individuals used to derive a statistic from a population
"n" represents the "sample size"

Random Sample

A select group of "units" from the population being described, chosen such that the process of selection is without bias and random

All members of the population must have an equal chance to be selected as a part of the sample

Population (N)

The total amount of individuals about whom the statistic aims to describe—it must be carefully defined!

Bias

A procedure/measurement is biased if the statistic offered by that procedure is not the same as the parameter

E.g., A statistic from a poll that, when following a strict procedure, is not representative of the population (voters).

-Say a poll was conducted without flaw, and Warren Litzke had 57% support going into election day

-She then receives 45%; we would say that statistic is biased

Margin of Error (MoE)

A measure of the uncertainty for a statistic or the estimate of a parameter.

Generally, We are 95% confident the true value, parameter, lies somewhere within the range of "stat +/- MoE"

-52% of our sample reports they will vote for Don ReSandis & our margin of error is roughly 3%

-Therefore, we are 95% confident that between 49-55% of voters will vote for ReSandis

Confidence Interval (90/95/99%)

A confidence interval is an interval constructed from data in such a way that the probability that the interval contains the true value of the parameter can be specified before the data are collected.

As we are more confident that our interval contains the desired parameter, our MoE must be increased to expand the range of the interval.

Public Opinion Poll, Types of Polls & Types of Sampling

Public Opinion Poll

A public opinion poll is a type of survey or inquiry designed to measure the public's views regarding a particular topic or series of topics. It is conducted by a pollster who asks a series of questions a cross section of the public and then extrapolates the results to the whole public. Opinion polls are used to predict election results or estimate public attitudes on issues.

Types of polls:
Benchmark polls
Opinion polls
Tracking Polls
Focus Group

Types of Sampling

When a poll is conducted, a cross section of the public must be selected. Sampling is the method by which this cross section is selected. There are various types of selection techniques and they are grouped into 2 categories (probability sampling and non-probability sampling).

In probability sampling methods, every element of the population has an equal chance of being selected. In non-probability sampling methods, all elements do not have an equal chance of being selected.

probability sampling methods
Simple random – Every individual in a group is chosen entirely at random and has the same probability of being selected.
Systematic/random – The first individual is selected randomly and subsequent individuals are selected using a fixed interval.

Stratified – The population is divided into several subgroups (such as gender or age) and then individuals are chosen randomly from these subgroups.
Cluster sampling – Subgroups of a population are chosen at random rather than individuals. After a subgroup is chosen, all individuals of the subgroup are then included in the sample.

non-probability sampling methods
Convenience – The individuals that volunteer for a poll are chosen for it.
Quota – Individuals are chosen for the poll based on certain characteristics.
Judgement/Selective – the judgement of experts is used to selected individuals for a poll.
Snowball - The first individual is selected randomly and they recommend further individuals to be selected for the study.

Polling Errors & Biases:

1) *The Most Common Error: Sampling Bias*

-Since the sample selection of any poll will determine the outcome of any such poll, bias in the selection of the sample will inevitably shift the outcome

- -Since it is impossible to have a perfectly balances sampling bias, a small degree of error can be expected
  - -However, the most egregious errors occur when an entire representative population that is supposed to be included in a sample is ignored, or a certain representative population is conspicuously overrepresented

-Famous example in recent times: 2016 Presidential Election

- -The general election polls, many of which were from highly respected polling agencies (Rasmussen, ABC, NBC, Fox News, Emerson, etc.) indicated strongly that Hillary Clinton would win the election, even having a double-digit lead over Donald Trump
- -What best explains this is a bias in sampling, that is, democratic voters were overrepresented in the polling samples

2) *Non-Response Bias*

-This kind of bias occurs when there is a failure of a certain representation of the sample to respond to pollsters and answer the survey

- -For example, if a certain population that buys a product 3 years before a company conducts a survey than a population that buys the product only 3 months before a survey

-The total amount of people that respond to the survey is not always the most salient factor

- -Even of all the people that respond to the poll bought a product or engaged with company close to the time that the survey is conducted, as long as those people are scrupulously chosen as accurately represented of the population, that can suffice for the accuracy and veracity of the poll

3) *Question Bias/Acquiescence Bias*

-This bias occurs when responders to a survey are more likely to say "yes" (or no) to a question because of the acquiescence of the question

-For example, is a survey asks a question such as "Do you support the congress taking action to improve the educational system in the U.S."?

-Obviously, this question is too vague to gage any valuable information about the sample's preferences

-Moreover, nearly everybody will say yes to a question like this, because in saying no ostensibly conveys that you are morally in the wrong or that you have poor intentions in answering the survey

-This type of bias is indicative of why the question must be articulated in such a way so that the acquiescence is less likely to occur and the stakeholders and responders will be able to answer the question honestly

4) *Timing Bias*

-Relates to any issues with the timing of the survey (ex. It is conducted too quickly, over too long of a time), typically occurs when the survey is conducted too soon or too late or over an extended period.

-Keep surveys within a tight window of time

-2-3 days > 5-6 days >>> multiple weeks

5) *Response Bias*

- Generally, mostly opinionated people will respond to a poll

-To avoid…

1. Make surveys short and easy
2. Keep the language simple and easily understandable by voters

-Primaries draw opinionated voters, so this bias should be watched.

-primary voter enthusiasm may make this bias a nonissue?!

Types of Polls

SMS/Online/Email

Text/SMS

2 Methods

- Text back and forth (Q->A->Q->A->Q->A).
- Send link to web survey.

Legal limitations

- FCC regulations make it illegal for companies to send text messages without expressed consent. (Question for legal team.)

Restricts question length (160 Characters).
Lower response rates than telephone surveys

- If used, keep response rate low.
- Lose interest more quickly than other types.

Survey mode	Questionnaire Length	Response rate	Breakoffs
SMS	5	12	8
SMS	12	13	19
SMS-web	5	12	1
SMS-web	12	11	10
Phone	5	38	0
Phone	12	41	0

Supplementary—not stand alone.

- Does not improve sample composition.

Online

Principle Source of data collected now—i.e. not phones.
Online polls do not work simply because of their design.

- demographics and other characteristics can predict vote choice.
- polls are weighted to match the country's demographics.
- WEIGHT > DESIGN

NOT traditional and weighting it is difficult.
Phone surveys more established

- Increasingly inaccurate
- E.g. Quinnipiac Polls

YouGov selects individuals from its panel of respondents.

- to match the demographic profile of American demographics.
- "synthetic sampling"
- Non-random

VoteCast used a mixed sample.

- Online + Phone
- Unadjusted Online poll was a mess.

Bogus Respondents:
https://www.pewresearch.org/methods/2020/02/18/assessing-the-risks-to-online-polls-from-bogus-respondents/

Email

Generally, the email survey approach takes one of three forms:

1. Software file attached to the email.
2. Electronic document attached to the email.
3. Questionnaire text embedded in the email itself.

Same coverage issues as Web surveys, in that not everyone has an email address.
Literature is not as well-developed or nearly as recent.
May be a good way to send out online survey links

Interactive Voice Response (IVR)

IVR is a method of polling that allows people to interact with a computer operated phone system using voice and DTMF tones input via a keypad.

How it works

The first step involves programming a computer with polling questions and possible answers. These computers then make automated calls to randomly selected phone lines. For example, a list of registered voters can be loaded into the computer and members of this list will be randomly selected and called. Many phone lines can be dialed at once and the most powerful systems can dial hundreds of phone lines at once.

Advantages

Ease of assembling a large sample
Low cost
Questions are asked identically each time

Disadvantages

Low response rate
Questionnaire must be short
Early hang ups

EPoliticalUSA and Custom IVR Solutions are 2 companies that specialize in polling for different organizations. An organization pays a fee to these companies, who in turn conduct polls according to the specifications given by the organization.

Additional notes:

According to the Harvard Business Review, IVR only works in primaries where the population is older and whiter. They suggest live polling is a better polling method.

Federal Law prohibits calling cell phones with IVR machines.

Live Calling

Traditional method of polling
Outdated, becoming obsolete
NYT Article: “The transition away from live interview telephone polling is well underway”
Live Polls are being replaced by Automated calls and online polls and all other modern methods

Advantages

- Real Time
  - Has the potential to provide more accurate results since the participant has less time to answer and has to give a more spontaneous answer
  - The surveyor has the opportunity make the participant more comfortable answering the survey, thus providing an answer more accurate to the participant’s true feelings
- Speed
  - The questions are asked right away, and answers are received right away, and then the results of the questionnaire are collected right after the call
  - Same speed or slower than automated calls
  - Higher quality experience
  - Easier to judge the quality of the poll than online methods

Disadvantages

- More expensive than modern digital polling because surveyors have to be hired
- Requires an institution that has the resources, namely the people and the money, to be able able to conduct Live Calls
- They will become more expensive as digital alternatives become the norm, especially if the quality is valued
- Not as easy to store the data as online polls which go directly into a database
- Fewer people are responding to telephone polls
-Slower

Polling Agencies

Rasmussen

Rasmussen uses automated polling technology to conduct their polls. They do this to ensure each respondent hears each question the same way. Random digit dialing (phone numbers are randomly selected) is employed to automatically sample the population.

For people with no landline phones, an online survey tool is used.

After surveying a population, the data is processed to make sure it reflects the overall population. For surveys of all adults, census bureau data determines the populations targets. For political surveys, the census bureau is the starting point and then population targets are determined by:

1) National trends
2) State's voting history
3) Recent polling

For Rasmussen's tracking polls, poll samples are weighed to reflect a specific mix of Democrats & Republicans so that polls are not biased due to sampling differences. The mix is determined by polling and previous elections.

Rasmussen uses strongly agree, somewhat agree, not sure, strongly disagree, and somewhat disagree in their polls. This may cause respondents to voice a minority opinion instead of "not sure". Rasmussen notes that, due to their automated polling, poll results show a higher level of disapproval than other live surveys. This may be due to respondent not willing to be harsh to live pollsters.

Quinnipiac

DESCRIPTION

"Among other high-quality polls, the Quinnipiac University Poll has been used to determine which presidential candidates made the Republican and Democratic primary debate stages in the 2016 and 2020 election cycles. The Poll is also ranked in the top 10% of the most accurate election polls by 538, a well-known polling and statistical analysis blog created by Nate Silver." –Quinnipiac University

"The Quinnipiac University Poll is considered the gold standard in the business, frequently lauded by USA Today and other national media organizations for its information and accuracy." –Asbury Park

CONTACT

Director of the Quinnipiac University Poll - Douglas Schwartz, PhD

Phone: 203-582-5201

Email: pollinginstitute@quinnipiac.edu

SAMPLE NUMBERS

Sample size (n) – 1,000 individuals
Sample – 18+ RV/LV
Confidence Level – 95%
Margin of Error - ~03.1%

METHODOLOGY

-Trained students and non-students conduct polls.
-Interviews/surveys are conducted via CATI system. CATI is an acronym that stands for Computer Assisted Telephone Interviewing system.
-Also do live polling
-Use a random dialing program to select participants who use both landline and mobile phones.
-Polls are conducted in multiple languages: namely, English and Spanish. The option of both is given to the sample.
-Roughly 1000 individuals comprise the sample of a given survey.
-The Survey is conducted over a period of 5-6 days wherein all participants are surveyed.

BIAS

-Many calls were horrible—especially this cycle. (See ME Senate, and State Presidential Races, & more: https://poll.qu.edu/.)
-Called 87% of races correctly in Aug 2016.
-Called 83% of races correctly in May 2020.
-Even lower now considering 2020 was a mess.
-Data is weighted demographically after the data is collected/interviews are conducted. (Could this be where bias is introduced?)
-Could young, diverse students conducting interviews put off Trump voters?

538 RATING STATS

-B+
-Bias: D +0.2%(likely higher after 2020)
-Simple Average Error: 4.6%

Insider Advantage

FiveThirtyEight:

-In 2010: Nate Silver on FiveThirtyEight rated IA as being the 2nd most unreliable polling firm
out of a total of 63 firms (62/63).
-Bad reputation from 2004 election and 2008
-In FiveThirtyEight's most recent poll rankings IA received a grade of B-
-They use IVR and Online polls
-One reason the quality of their polling may be criticized is because they have stopped doing
live calls completely
-High simple average error of 7.2%, one of the highest if not the highest of the major polling
firms
-70% races called correctly
-Predictive percentage: 0.4% which is bad, negative scores indicate high quality
-Slight democrat leaning bias

Notes:

Son of the CEO of IA says that their online sample combines interviews from an unspecified online panel and respondents on Facebook
-Issues with weighting?
-Over multiple iterations (preserving original ratios), the online and telephone polls were integrated and subsequently resampled randomly. This poll was weighted for age, gender, and political affiliation" - Matt Towery Jr.
-By randomly resampling, you jettison all probability sampling, which has been criticized, because probability sample is the only technique that can be generalized to
-Have also been criticized for discarding data, waste of time, money, and energy
-They online select a portion of the results to be used in the final result (½)
-Done by placing advertisements on FB
-2:1 ratio of telephone to online surveys based on experimentation (trial and error)
-Online interviews are all registered voters

Trafalgar

Methods of Polling:

- Live Callers
- Integrated voice response
- Text messages
- Emails
- Private proprietary digital methods

-Nine questions at a time
-Based on perceptions about attention span (likely to avoid response fatigue)
-Designed to (quickly get opinions from those that don't usually participate in political polls)
-Intended to account for "Social Desirability bias", tried to gage the poll participants' true feelings where some would be reluctant to answer

*They typically refrain by asking strict questions like "who will you for?" and opt for questions that avoid social desirability bias like "who do you plan on voting for as of now?" or "who are you leaning towards voting for?"

-They only conduct smaller, state level polls
-They try to take "shy Trump voters" into account
-They have received a grade of C- by FiveThirtyEight
-Criticisms about lack of transparency about methodology
-They don't reveal much information about their polling methodology
-Emphasizes that the typical methodology of "polling industry" is outdated and not in line with modern times

Polling Reports:

- Georgia senate runoff
  - Conducted over the span of 2 days to avoid timing bias
  - 1056 (1000 typically thought of to be the correct range)
  - Only a 1.41% response rate (Low, but expected)
  - 2.94% margin of error (normal)
  - 95% confidence

Success:

- Only poll that accurately predicted that Trump would win Michigan in 2016
- Predicted that Trump would win the 2020 election in all of the swing states that Biden
  flipped and won

Data Vis: Grafana

Grafana

Used by Paypal, eBay, staples, Intel, Booking.com, Vimeo, OpenTable, etc

Open Source

Pricing:
- Free access to regular version
- Grafana Pro is $49 per month (I don't think we'll need it)

Products:
- Cloud storage
    - Only feasible option
- Enterprise
    - More Secure
    - You can only request a demo as of now,

Features:
- Abundant visualization options
    - Pie Charts, Bar Charts, Line Graphs (smooth ones too), scatter plot, etc
    - Dashboards with templates
    - Different metrics for "ad hoc queries"
    - Logs
        - Easy to switch between using metrics to logs
        - While logs are about a specific event, metrics are a measurement at a point in time for the system
    - Set maximums on certain data such that if the data exceeds that maximum, you will get an alert
    - Multiple data sources on one graph
    - Annotate graphs
    - You can filter your data (makes it more specific)

Supported Plugins:

- Used for collecting the data and delivering it to Grafana
- Plug-ins:
    - Snowflake
    - Github
    - Redis
    - Splunk
- Data sources:
    - Datadog
    - Jira
    - Google sheets (probably a no-go)
- Apps
    - Kubernetes
    - Zabbix
    - Synthetic Monitoring
- Visualization Panels
    - Pie Chart
    - Gantt
    - World map panel

-You can create documents locally without a plugin
-Sharing Dashboards is easy
-Plugins can be created (the Supported plugins are unnecessary but easier)

-Thousands of official and community made dashboards
-Dashboards are made through a Program called Prometheus

Data types:
Int, double, string, etc

https://grafana.com/docs/grafana/latest/packages_api/data/